

Project 2 - Data Driven Prediction Models of Energy Use in Appliances

1. Description

In this project, you must reproduce some of the results in the attached paper "*Data Driven Prediction Models of Energy Use of Appliances in a low energy house*" by Luis M. Candanedo et.al".

- 1) The dataset to be used is provided by the excel file "**energydata_complete.csv**"
- 2) You will train multiple regression estimator to make predictions. I suggest modifying the jupyter notebook given in lab 2 and carry out the same analysis.
- 3) In addition, before applying regression apply first PCA to reduce the dimension of the original data and observe whether it results speeding up the regression model.
- 4) Write a report in latex with the results and your observations

2. Data set

The data set records the energy consumption of appliances in a house along with the corresponding environment, weather and time statistics in a time span of 137 days. There are a total of 19735 pieces of data and 32 variables in the data set. A description of each variable is presented in Table 1.

Data variables and description.

Data variables	Units	Number of features
Appliances energy consumption	Wh	1
Light energy consumption	Wh	2
T1, Temperature in kitchen area	°C	3
RH1, Humidity in kitchen area	%	4
T2, Temperature in living room area	°C	5
RH2, Humidity in living room area	%	6
T3, Temperature in laundry room area	°C	7
RH3, Humidity in laundry room area	%	8
T4, Temperature in office room	°C	9
RH4, Humidity in office room	%	10
T5, Temperature in bathroom	°C	11
RH5, Humidity in bathroom	%	12
T6, Temperature outside the building (north side)	°C	13
RH6, Humidity outside the building (north side)	%	14
T7, Temperature in ironing room	°C	15
RH7, Humidity in ironing room	%	16
T8, Temperature in teenager room 2	°C	17
RH8, Humidity in teenager room 2	%	18
T9, Temperature in parents room	°C	19
RH9, Humidity in parents room	%	20
To, Temperature outside (from Chièvres weather station)	°C	21
Pressure (from Chièvres weather station)	mm Hg	22
RHo, Humidity outside (from Chièvres weather station)	%	23
Windspeed (from Chièvres weather station)	m/s	24
Visibility (from Chièvres weather station)	km	25
Tdewpoint (from Chièvres weather station)	°C	26
Random Variable 1 (RV.1)	Non dimensional	27
Random Variable 2 (RV.2)	Non dimensional	28
Number of seconds from midnight (NSM)	s	29
Week status (weekend (0) or a weekday (1))	Factor/categorical	30
Day of week (Monday, Tuesday. . . Sunday)	Factor/categorical	31
Date time stamp	year-month-day hour:min:s	–

Table 1

The energy consumption of appliances is recorded every 10 minutes in order to capture the quickly changing data. The energy consumption of lights and relative humidity recording are also used to predict whether a room is occupied. To study the impact of the weather condition, data from the nearest weather station is included in the data set.

The data set is split into two parts: the training data and the testing data. 75% of the data is used for training models and the rest is used for testing.

3. *References*

[1] Luis M. Candanedo. *Data Driven Prediction Models of Energy Use of Appliances in a low energy house*. Energy and Buildings 140 (2017) 81–9